AD-A100 262 NAVAL HEALTH RESEARCH CENTER SAN DIEGO CA F/G 5/10
SURVEY QUESTIONNAIRES: MORE THAN MEETS THE EYE.(U)
DEC 77 J M LAROCCO, M C BUTLER
UNCLASSIFIED NAVHLTHRSCHC-77-57 NL

END
DATE FILMED
7-81
DTIC

LEVEL II

(2)

SURVEY QUESTIONNAIRES: MORE THAN MEETS THE EYE

J. M. LA ROCCO

M. C. BUTLER

REPORT NO. 77-57

NAVAL HEALTH RESEARCH CENTER
P. O. BOX 85122
SAN DIEGO, CALIFORNIA 92138

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

Survey Questionnaires:

More than Meets the Eye*

James M. LaRocco and Mark C. Butler

Naval Health Research Center
San Diego, California 92152

During the past several years, there has been a growing resentment throughout the fleet toward both the number and type of survey questionnaires imposed upon operational naval units. Perhaps the most frequent criticism is that such surveys are unjustified because the results are often equivocal or misleading and fail to lead to any noticeable changes in policy or practice. Furthermore, it has been argued that such surveys are intrusive in that they seek too much information from respondents and place an unnecessary burden on already overworked personnel. To members of the Navy's professional research community, these and other problems associated with large-scale survey research have become a matter of paramount concern.

---

*Report Number 77-57, supported by Naval Medical Research and Development Command, Department of the Navy, under Research Work Unit ZM51.524.022-0007. The views presented in this paper are those of the authors. No endorsement by the Department of the Navy has been given nor should be inferred.

In an attempt to address these objections and criticisms, it may be of value to explain the rationale that underlies questionnaire development. In other words, if more people were aware of the nature of survey questionnaires and could ascertain whether a proposed instrument met certain specified standards, questionnaires would be less objectionable and could conceivably yield higher quality information. In addition, with some knowledge of what constitutes a good questionnaire manpower managers could better assess the potential usefulness of survey results. Unfortunately, many survey questionnaires currently in use lack a sound basis in theory and are methodologically inadequate to address their stated purposes. Thus, results generated by these instruments are frequently unclear or misleading.

Often unknown or neglected *is the fact that questionnaire development is* a technical specialty requiring considerable training in such areas as psychology, mathematics, and some aspects of computer science. The combination of skills listed above defines the branch of psychology called psychometrics. Psychometrics deals specifically with the development and application of mathematical procedures to construction of survey questionnaires, aptitude measures, ability tests, and so forth. The training required to gain more than a superficial understanding of basic psychometric principles can be described as rigorous; one wonders what percentage of surveys currently used in the Navy have been subjected to such rigor.

The issues discussed briefly above regarding test and survey development and interpretation are not new for health and behavioral researchers. In fact,

such concerns prompted the American Psychological Association (APA) to publish a set of guidelines that outline professional and ethical standards for constructing and using educational and psychological tests. (In the current discussion, the term "test" refers to any instrument that elicits evaluative responses, including opinion or attitude surveys.) The APA Standards Manual makes one point very clear: In spite of popularized notions and common practice, one cannot simply devise a series of questionnaire items based on job-related experience, label such questions "job satisfaction" or "task analysis," and expect to accurately measure those concepts in either a meaningful or interpretable way. Although such questions often possess what is known as "face" validity, as the APA emphasizes in their official Standards of Educational and Psychological Tests,[1] "...so-called 'face' validity, the mere appearance of validity, is not an acceptable basis for interpretive inference from test scores" (p. 26).

This is not to say that job-related experience cannot be an important adjunct to questionnaire development. On the contrary, such experience can be invaluable in assuring that job characteristics and other aspects important to the individual and the organization are explored and appropriate terminology is used. Thus, questionnaire development may require a team approach with line personnel supplying information about content areas to be explored and a staff of technical specialists providing the expertise to convert that information into a viable instrument. This is standard procedure for the development of any piece of equipment to be used by the Navy, and a survey instrument should be no exception. Unfortunately, the majority of personnel who must respond to

questionnaires are not part of, and are generally unaware of, the extensive efforts required in survey development. The present article will attempt to impress upon the reader that certain prerequisites must be met to produce an acceptable survey instrument. In the remainder of this paper, the basic characteristics that a survey questionnaire should possess to provide truly useful information are described.

To be useful, a questionnaire must have the two psychometric properties of reliability and validity. Briefly stated, reliability is the degree to which items, groups of items (i.e., scales), and the test itself yield consistent results over time. Thus, the essence of reliability can be expressed as consistency of measurement. More specifically, reliability has been traditionally conceived of as the degree to which scores derived from one test administration will resemble scores derived from a second administration of the same test to the same individuals. If the pattern of test-retest scores is highly similar, then the particular test instrument may be viewed as "reliable," i.e., it possesses stability over time. Another and in some ways more common method of estimating reliability is to ascertain the extent to which a given set of items tap a common domain. Thus, the degree of interrelatedness or internal consistency of the item is taken as an index of reliability.[2]

The implications and importance of reliability to theory construction based upon the two approaches to reliability (i.e., temporal stability and internal consistency) are too complex to be dealt with in detail here. Briefly, however, an investigator needs to be confident that the instruments employed in a particular study will adequately assess the topic of interest

over multiple measurement opportunities. Such confidence is not acquired easily; it requires many long hours of writing and refining items and evaluating individual responses. For example, a question should be written so that it can only be interpreted in one way. Considering the different educational levels, experiences, and backgrounds of test takers, this is not so easy as it sounds. Even the response choices must be considered carefully. Should there be four or five possible choices? Perhaps three would be adequate. What is the difference if one choice is "occasionally" and another is "not usually"? These and myriad other details must be weighed, for each is a potential source of unwanted error and, hence, can reduce reliability. Without such detailed efforts, however, confidence in the stability of the survey instrument suffers, and even the meaning and usefulness of the information received must be questioned.

Similarly, with respect to estimates of internal consistency, the clarity, precision, and accuracy of the survey scales (i.e., groups of items) or items themselves receive primary consideration. In short, if the researcher is interested in relating attitudes about certain aspects of the individual's work setting (i.e., job satisfaction, motivation, leadership) to retention in the Navy, a high degree of confidence should be placed in the fact that the various scales used to measure job-related attitudes in reality clearly pertain to the specified dimensions of the work environment. Accomplishing this end requires that each scale representing a specific dimension (e.g., satisfaction with pay) contain enough items to measure that facet. In addition, several scales are often required to tap a general domain (e.g., job satis-

faction). Thus, the price one pays for a psychometrically sound instrument is reflected in part by the length of the questionnaire and the apparent redundancy of items. Unfortunately, these latter concerns (length and redundancy) constitute the major source of complaint regarding questionnaires.

Once a test has been constructed and determined to be reliable, its worth as a measurement instrument still must be demonstrated further. Research designed to evaluate the usefulness of tests, scales, or survey questionnaires has as its foremost challenge the demonstration of the validity of the findings produced by the use of the particular instrument.

Validity is a mathematically determined index that allows us to reach conclusions about how faithfully a questionnaire or test represents some domain of interest.[3] There are several approaches to validity, including content validity, criterion-related validity, and construct validity. Content validity refers to the problem of determining whether or not the content of a scale or test under consideration adequately represents the dimension being measured (e.g., satisfaction with pay, leadership, etc.). It is at this point that the years of experience reflected by members of the line community can be meaningfully interfaced with the technical skills of the research scientist. In other words, items that convey meaning and content to the line community because of that community's experience and knowledge also are likely to be effective indices of the content in a test. Thus, the researcher should solicit evaluative responses regarding a particular issue from a variety of knowledgeable persons. The responses then would be analyzed and those that demonstrated both appropriate and acceptable psychometric properties would be

selected and included in the final instrument.

The second type of validity, criterion-related, is perhaps the most relevant to Navy managers. Criterion-related validities apply when one wishes to use the test score to infer an individual's standing on some other variable--the criterion. Examples of potential criteria are reenlistment decisions, fitness reports, grades, scores on battle problems, and so forth. The criterion-related validity of greatest importance is predictive validity--the extent to which an individual's future level on the criterion can be predicted from a knowledge of prior test performance. For example, the usefulness of college entrance exams is predicated on their ability to predict successful completion of four years of college. The magnitude of this relationship to collegiate performance is referred to as predictive validity. In an example closer to home, research on retention has shown that a weighted average of measures of pre-service anti-social behavior (arrests, school expulsions), educational level, and age is a valid predictor of completing a four-year enlistment and being recommended for reenlistment.[4]

Related to predictive validity is concurrent validity--the extent to which a score may be used to estimate an individual's present standing on the criterion. Sometimes concurrent validity is used to infer predictive validity. For instance, when determining what variables might be related to reenlistment it would be necessary to administer a questionnaire and then wait several years until all respondents have had the opportunity to reenlist. This is costly in terms of time and money, so questionnaires often contain an item asking the individual's intent to reenlist. Several studies which followed men through

their enlistments have shown that intent to reenlist measured only six months after enlistment is highly related to actual reenlistment (i.e., it is a valid predictor). Therefore, to save time and money many studies use intent to reenlist as a criterion rather than actual reenlistment. For instance, the relationship between job satisfaction and intent to reenlist (concurrent validity) is used to infer the predictive validity of job satisfaction with respect to actual reenlistment behavior. While this is a common practice even among knowledgeable researchers, concurrent validation should be recognized for what it is--a method to estimate the probable magnitude of a potential predictor-criterion relationship as it may appear at some point in the future.

It should be clear that the major value of criterion-related validities is in the area of applied research where the basic question is one of determining both the extent of particular problems and the most effective means of addressing those problems. Thus, it is often criterion-related validity that is most relevant to the interests of management personnel. On the other hand, the most important aspect of the validation process for theoretical research is referred to as <u>construct validity</u>. The importance of various construct validation procedures emerges more clearly when one considers that such procedures are specifically concerned with the establishment of relationships between actual data (either direct behavioral observation or questionnaire responses) and hypothesized concepts or constructs. Such constructs are the attributes, beliefs, individual characteristics, and personality traits inferred from psychological research upon which the foundation of theoretical development rests. Specific details regarding construct validation procedures

are not relevant to the current discussion, but suffice it to say that the accumulation of content, predictive, and concurrent validity information almost invariably leads to construct validation and, hence, scientific (i.e., theoretical) advancement.[5] It should be emphasized that this final goal, scientific advancement, cannot be achieved without strict adherence to the principles of theory and test development.

In this brief note, only the most basic issues in test construction have been presented. The purpose of this presentation has been to inform the reader that the development of a properly designed survey instrument is based upon logical and defensible mathematical properties and upon well-established principles regarding the measurement of human attributes and abilities (psychometrics). It is only because of this rigorous foundation that management personnel can accept the information such surveys provide, assess the degree of confidence to be placed in the results, and ultimately apply the findings to the everyday problems of the Navy. The corollary of this should be obvious; those survey instruments that lack a foundation of theoretical and methodological rigor can only serve to increase fleet-wide problems because of the unreliable and invalid information they produce.

The reaction to the large number of surveys administered throughout the Navy has led to a cry for an unconditional halt to such research aboard naval units. But problems of absenteeism, desertion, low retention, and other indicators of personnel dissatisfaction and poor performance are still with us and are, in fact, reaching alarming proportions. If the management steps necessary to reverse these trends are to be on a sound basis, research must con-

tinue. In the words of Admiral Smedberg:[6] "The performance measurement of large Navy systems requires measurement in the operational environment, either at a shore establishment or at sea, because it is impossible to put such large systems (e.g., a ship) in a laboratory.... Although measurement in the operational environment creates special difficulties for the investigator, not least the need not to interfere with on-going operations, it seems to me that we cannot achieve the Navy's performance measurement desires unless measurement is rooted in that environment" (p. 10).

At the same time, something must be done to insure that when personnel time and effort is required, it is not unreasonable to expect that the results obtained will merit the man-hours expended, both at the fleet and higher policy-making levels. Rather than eliminating surveys altogether or basing approval for a particular survey administration on either the contents of the questionnaire or the purported purpose of the study, it makes more sense to require from the individual or organization soliciting the approval theoretical and empirical justification for their proposed work. In addition, the applicability of the proposed work toward solving the problems of the Navy should be made explicit, that is, the responses obtained from the participating individuals should reasonably be expected to provide answers to the research questions posed. The ultimate criterion for implementing a survey, then, should not be that the survey will provide answers but that the answers provided by the survey will be dependable (i.e., reliable) and would actually address the problems or domains of interest (i.e., be valid).

Footnotes

[1]American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (Joint Committee). Standards for Educational and Psychological Tests and Manuals (Washington, D.C.: American Psychological Association, 1974).

[2]It is important to note that internal consistency reliability procedures merely estimate the reliability of a test based upon the degree of interrelatedness among the constituent items. For a more thorough and detailed discussion of all reliability techniques, the reader is referred to Thorndike, R. L. "Reliability," as published in D. N. Jackson and S. Messick (ed.), Problems in Human Assessment (New York: McGraw-Hill Book Co., 1967).

[3]Ibid., Note 1, p. 25.

[4]LaRocco, J. M., Pugh, W. M., and Gunderson, E. K. E. "Identifying Determinants of Retention Decisions," Personnel Psychology (vol. 30, 1977), pp. 199-215.

[5]In a technical sense all validities are forms of construct validity. The interested reader can find a discussion of this concept in Guion, R. M. "Open a New Window: Validities and Values in Psychological Measurement," American Psychologist (vol. 29, 1974), pp. 287-296.

[6]Smedberg, W. R. "The Navy's Need for Performance Measurement," a paper presented to the Navy Personnel Research and Development Center Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems (San Diego, Calif., October 1977).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 77-57	2. GOVT ACCESSION NO.: AD-A100262 (4)
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): (6) Survey Questionnaires: More than Meets the Eye.	5. TYPE OF REPORT & PERIOD COVERED: Final rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (5) James M. LaRocco and Mark C. Butler	8. CONTRACT OR GRANT NUMBER(s): (12) (13)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Health Research Center, P.O. Box 85122, San Diego, CA 92138 (17)	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: ZM51 524 022-0007
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Medical Research & Development Command (11), National Naval Medical Center, Bethesda, MD 20014	12. REPORT DATE: December 1977
	13. NUMBER OF PAGES: 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Bureau of Medicine and Surgery, Department of the Navy, Washington, DC 20372 (16) ZM51524	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(14) NAVHLTHRSCHC-77-57

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reliability
Validity
Survey
Test construction
Naval personnel

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In an attempt to address some objections and criticisms by line personnel concerning survey questionnaires, issues basic to test and survey development were discussed in a nontechnical manner. Three major points were made. First, questionnaire construction requires extensive training in psychology, mathematics, and some aspects of computer science. Second, prior to judging the applicability of survey results, one must ascertain the reliability and validity of the measures involved. Finally, the authors specify that implementation of a survey should be based on evidence provided by the researchers that their questionnaires are psychometrically sound.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)
Unclassified

391642